Accred Qual Assur (2002) 7:378–387
DOI 10.1007/s00769-002-0531-6

Elisabete A. De Nadai Fernandes
Fábio S. Tagliaferro
Adriano Azevedo-Filho
Peter Bode

Organic coffee discrimination with INAA and data mining/KDD techniques: new perspectives for coffee trade

Received: 12 April 2002
Accepted: 31 July 2002

E.A. De Nadai Fernandes · F.S. Tagliaferro
Nuclear Energy Center
for Agriculture (CENA),
University of São Paulo (USP),
PO Box 96, 13400–970 Piracicaba,
São Paulo – Brazil
e-mail: lis@cena.usp.br
Tel.: +55-19-34294655
Fax: +55-19-34294654

A. Azevedo-Filho
Department of Economics
and Centre for Advanced Studies
in Applied Economics,
College of Agricultural Engineering Luiz
de Queiroz (ESALQ),
University of São Paulo (USP),
PO Box 9, 13418–900 Piracicaba,
São Paulo – Brazil

P. Bode
Interfaculty Reactor Institute (IRI),
Delft University of Technology (TUDelft),
Mekelweg 15, 2629JB Delft,
The Netherlands

Abstract Samples of green coffee (*Coffea arabica*) produced in the crop year 1999/2000 in Minas Gerais state, Brazil, were analyzed for the elements Al, Ba, Br, Ca, Cl, Co, Cs, Cu, Fe, K, Mg, Mn, Na, Rb, S, Sc, and Zn using instrumental neutron activation analysis (INAA), in an attempt to establish fingerprints of organically grown coffee. Using data mining/KDD techniques the elements Br, Ca, Cs, Co, Mn, and Rb were found to be suited as markers for discrimination of organic from conventional coffees.

Keywords Organic coffee · Inorganic composition · Discrimination methods · Pattern recognition · Quality demonstration vs. quality designation · Data mining · KDD

Introduction

During recent years, there has been an increasing concern among consumers about new coffee quality attributes associated with the absence of chemical contaminants, negative environmental impacts caused by the production system and use of bad labor practices. This concern is addressed, at least in part, by production methods identified with the concepts organic, sustainable, ecological and biological. Because of this trend, traditional coffee producers are increasing the supply of organically cultivated coffees to meet the growing demand of markets such as the EU, Japan, and the USA. Importers, however, are facing numerous problems for the discrimination of organic from conventional coffees in order to discover and to avoid frauds.

The availability of reliable certification procedures for organic coffee might facilitate and strengthen the international trade of this product. Current certification procedures rely, strongly, on process certification instead of on product certification. The organic attribute is, therefore, often designated to a coffee lot by an attached certificate rather than demonstrated by objective procedures. Because higher international prices are achieved by organic coffee and certification relies on quality designation, there might be an incentive in the market to sell conventional coffees or mixed coffees as being pure organic. This problem would be minimized if the intrinsic quality of the product could be objectively demonstrated by means of fingerprints allowing its correct identification. The idea of quality demonstration developed in this research was inspired in measurement competence certification based on demonstration as suggested by De Bièvre and Taylor [1].

Several attempts have been devoted to either establish the regional origin of coffee [2] or to differentiate between Arabica and Robusta varieties, both in green and roasted coffees using elemental composition [2–4] and chemical attributes (chlorogenic acid, caffeine, trigonelline, aqueous extract, amino acids, and polyphenols) [5]. The aroma fraction can successfully be employed to characterize roasted coffees of different origins [6].

Appropriate identification of organic coffee means protection to producers and consumers, as well as new perspectives for international trade. It may be based on the determination of the agrochemical compounds and residues in the coffee beans, which is to some extent troublesome due to the high costs and scarce availability of the associated standards. Moreover, pesticides used in coffee plants can be transformed during the roasting process of coffee beans leading to misidentification.

This study provides an assessment on whether elemental composition, measured by instrumental neutron activation analysis (INAA), can discriminate organic from conventional coffees. INAA is an advantageous technique for this kind of study because of its multielement character, absence of a dissolution step and no need for matrix-matching multielement standard [7]. The analysis on the perspectives for discrimination considered samples of organic and conventional coffees from a major production region in the Minas Gerais state, Brazil. The information from such samples was organized in a database and explored by a data mining/KDD approach [8, 9] with the objective of looking for fingerprints, which would allow the discrimination. The next section provides some background information on organic agriculture and its definitions.

So what does organic mean?

Organic agriculture is growing worldwide. Currently (2002) it has been practiced in more than 120 countries. The total area certified as organic encompasses 20 million hectares and the market of organic products is responsible for USD 20 billion. Italy has around 50,000 certified organic farmers, the largest number in a single country, while Australia has the largest area covered with 7.7 million certified organic hectares. Developing countries also have a significant participation in the sector accounting for hundreds of thousands of farmers practicing organic farming [10].

In spite of the increasing interest for organic agriculture, there is no single definition accepted worldwide for this type of agriculture. The International Federation of Organic Agriculture Movements (IFOAM) defines organic agriculture as “all agricultural systems that promote the environmentally, socially and economically sound production of food and fibers” [11]. The joint FAO/WHO Codex Alimentarius Commission, in the Guidelines for the production, processing, labelling and marketing of organically produced food (GL 32, 1999) defines organic agriculture as “a holistic production management system that promotes and enhances agro ecosystem health, including biodiversity, biological cycles, and soil biological activity” [12]. These general definitions are formalized in technical guidelines/standards that specify the allowed practices for organic agriculture [11, 12].

In some countries, organic agriculture is also characterized as ecological or biological agriculture. All these traditions were taken into account by the European Commission when drafting the EU-Regulation 2092/91, protecting the use of all three terms – organic, biological and ecological – including abbreviations like bio or eco in the EU official languages [10].

Experimental

Sample collection

Samples were collected in Santo Antonio do Amparo, Minas Gerais state, one of the pioneer cities in the production of organic coffee in Brazil. Five Arabica coffee fields cultivated under different systems were selected: two organic, two conventional and one in-conversion, which is a field changing from conventional to organic system in a timeframe of at least 5 years. The organic coffees were produced in accordance with the guidelines from the Instituto Biodinâmico (IBD), an IFOAM accredited member, and the Associação de Agricultura Orgânica (AAO), a Brazilian affiliate of IFOAM [11]. After harvesting, the coffee beans were processed and stored in heavy fabric bags (60 kg), inside an ambient-controlled warehouse, composing individual batches for each harvested field. Then, 25 samples of approximately 0.5 kg were taken from each batch, totaling 125 samples.

Sample preparation

Sample preparation was carried out at the Radioisotopes Laboratory, CENA/USP, Piracicaba, Brazil. Samples were oven-dried at 60 °C until constant weight, ground in alumina mill and test portions of 500 mg placed in special polyethylene capsules for irradiation (Vrije Universiteit Amsterdam). The moisture content was assessed by replicates taken during capsules filling. Certified reference materials (NIST/SRM 1515 – Apple Leaves and NIST/RM 8433 – Corn Bran) were used for internal quality control.

Instrumental neutron activation analysis (INAA)

The samples were irradiated using the facilities of the 2 MW nuclear research reactor “Hoger Onderwijs Reactor” of the Interfaculty Reactor Institute, Delft University of Technology. The INAA Laboratory at this institute operates with an accredited quality system since February 1993 (original accreditation by STERLAB for compliance with EN45001; nowadays by the Dutch Council for Accreditation for compliance with ISO/IEC 17025).

Two irradiations and three measurements were performed for the multielement determinations. First, elements based on short-lived radionuclides were determined by irradiation for 30 s in the fast rabbit system, under a thermal neutron flux of approximately 1.6×10^{13} cm s^{-1}. After 1 min decay time, the induced activity was measured for 5 min using a Ge detector (12% relative efficiency) at a sample–detector distance of 5 cm. Metallic zinc foils irradiated together with the samples allowed to estimate the neutron flux; the activity of the foil was measured after the measurement of the sample. For the determination of elements based on medium and long-lived radionuclides, samples were irradiated 4 h under a neutron flux of approximately 5×10^{12} cm ^{2} s^{-1}. Each irradiation consisted of a batch of 14 samples, neutron flux monitors and internal quality control samples. After 4–5 d decay time the induced activities were measured on a Ge detector (17% relative efficiency) for 1 h (sample–detector distance 5 cm), and after approximately 21 d in a well-type Ge detector for 1 h. The neutron flux was again estimated using Zn monitors. The elemental quantification is based on the single comparator method. The element calibration constants in this method have all been previously experimentally established in the same irradiation and counting facilities, using working standards made from element compounds of known composition. Correction factors are being applied if changes in the neutron spectrum occur due to alterations of the reactor core configuration. More details on the operation and quality assurance in this laboratory can be found elsewhere [13, 14].

Data mining/KDD methodology

Data mining is a concept used by practitioners of artificial intelligence, computer science, and statistics to indicate the process of knowledge discovery in databases (KDD). Data mining or KDD relies on multidimensional data visualization techniques, machine learning and pattern recognition methods, as well as on standard statistical methods to perform the nontrivial extraction of implicit, previously unknown, and potentially useful information from data [8, 9].

A formal representation for the discrimination/pattern recognition problem investigated in this research is presented in the following paragraphs.

Let S be a coffee sample belonging to an unknown category of interest (organic, conventional and in-conversion); CAT_i be a category i from Ω_{CAT}, a set of categories of interest; $\mathbf{E(S)}$ be a vector (size n) of features (in this case elemental concentrations) in sample S; $R(\mathbf{E(S)})$ be a (multivariate) $n\times m$ function of the features (elemental concentrations); N_j be a particular subset j of the R^m space (m-dimensional space), associated with function R and vector of features $\mathbf{E}$. This subset j is an element of Ω_N, the set of possible subsets.

A first issue of interest here is whether there would be an appropriate definition of a function $R(\mathbf{E(S)})$ of the elemental concentrations that would lead to meaningful information on the probabilities of membership of this coffee sample S in a specific category CAT_i, given the observation that $R(\mathbf{E(S)})\in N_j$, that is, the function of the elemental concentrations presented a value within the category of values N_j, or, algebraically:

$$\Pr\left[S\in \mathrm{CAT}_i \middle| R(\mathbf{E(S)})\in N_j\right],\quad \mathrm{CAT}_i\in\Omega_{\mathrm{CAT}} \tag{1}$$

Equation (1) can be represented (under Bayes rule) by:

$$\frac{\Pr\left[R(\mathbf{E(S)})\in N_j \middle| S\in \mathrm{CAT}_i\right]\Pr\left[S\in \mathrm{CAT}_i\right]}{\sum\limits_{\mathrm{CAT}_i\in\Omega_{\mathrm{CAT}}}\Pr\left[R(\mathbf{E(S)})\in N_j \middle| S\in \mathrm{CAT}_i\right]\Pr\left[S\in \mathrm{CAT}_i\right]},\quad \mathrm{CAT}_i\in\Omega_{\mathrm{CAT}} \tag{2}$$

In Eq. (2) the last term in the numerator represents the *a priori* probability of having the coffee sample S classified in CAT_i without any information on elemental concentrations. A comprehensive database with information for each coffee sample, including elemental concentrations and cultivation system category, might facilitate the estimation of the conditional probabilities in Eq. (2).

A second issue of interest is the perspective that there might be more than one function $R(\mathbf{E(S)})$ of the elemental concentrations able to provide useful information under the framework discussed so far. It is possible, for instance, that the knowledge of elemental concentrations from two different subsets of elements, {Na, Rb, Sc} and {Br, Fe}, have comparable power in discriminating the cultivation system. In this situation, cost-effectiveness issues could be taken into consideration to facilitate the selection.

A number of parametric/nonparametric approaches could be potentially helpful to provide insights into the pattern recognition/discrimination problem formalized in the previous paragraphs. Some of these approaches (such as graphical analysis) do not provide quantitative answers but contribute to a better understanding of this problem without unnecessary complications.

In this research, graphical analysis as well as nonparametric methods (mostly), were used to investigate the feasibility and nature of cost-effective solutions which might lead, with further refinement, to routine methods for discrimination of coffee attributes or categories (organic, conventional, in-conversion), based on elemental concentrations. These techniques include multivariate data visualization methods, classification trees for categorical response, and data driven Bayesian networks. Nonparametric techniques were used because they are often more robust to outliers, less dependent on strong distributional assumptions, and allow a more straightforward treatment of missing values/incomplete information, as well as nonlinearity, than their parametric counterparts.

The implementation of KDD methods required the organization of the available information on coffee samples from known categories in a database – generally speaking, a table readable in digital format. This table includes in each line the available information from each sample, including cultivation system category, soil contamination category and elemental concentrations, organized in each column table.

Results and discussion

The mean concentrations and coefficients of variation of Al, Ba, Br, Ca, Cl, Co, Cs, Cu, Fe, K, Mg, Mn, Na, Rb, S, Sc, and Zn for the five cultivation systems/soil contamination categories studied are shown in Table 1.

A preliminary analysis considered an R implementation [15–17] of univariate data visualization methods (with box plots) to explore features of the elemental concentrations for each category of interest (C, conventional coffee; O, organic coffee; T, transition or in-conversion). A typical box plot presents the data between “hinges” roughly associated with the quartiles 1, 2, 3, and 4, with the central hinge being the median. Samples suggested as outliers are plotted as circles, off the quartile ranges.

Figure 1 presents box plots for elemental concentrations found in the database of samples, for each category of interest. The organic (O) category aggregates the samples from the two organic sources available. The conventional (C) category includes samples with and without soil contamination. Concentrations of Br, Ca, Mn, and Rb for organic samples tended to be substantially lower or higher than those for other samples. Lower concentrations of Cs were observed for category in-conversion

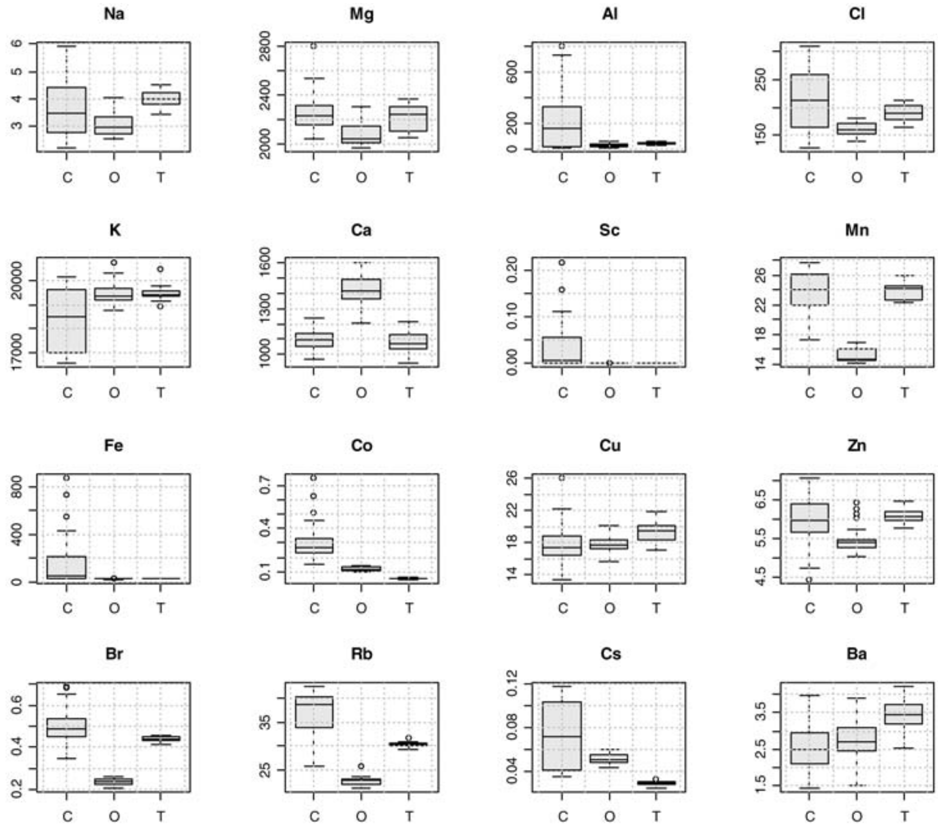


Fig. 1 Box plots of elemental concentrations (μg g^{-1}) for each category (includes the soil-contaminated samples in the conventional category)

Table 1 Mean elemental concentrations (μg g^{-1}) of coffee beans

Category		Al	Ba	Br	Ca	Cl	Co	Cs	Cu	Fe	K	Mg	Mn	Na	Rb	S	Sc	Zn
Conventional	Mean	31	2.3	0.45	1120	157	0.21	0.10	17	27	17,200	2160	25	2.8	33	1510	0.00086	6.1
	CV (%)	62	22	11	5	9	20	8	6	4	3	4	15	12	12	14	27	13
Conventional+soil	Mean	340	2.8	0.55	1070	256	0.37	0.042	19	276	19,600	2330	23	4.6	40	1950	0.063	5.8
	CV (%)	53	21	12	5	8	29	12	12	73	2	5	7	11	2	22	76	6
In-conversion	Mean	48	3.5	0.44	1080	191	0.054	0.029	19	32	19,400	2200	24	4.0	30	1580	0.00067	6.0
	CV (%)	22	12	3	6	7	6	6	6	3	2	5	5	7	2	21	14	3
Organic 1	Mean	39	2.8	0.22	1480	166	0.14	0.049	18	26	19,600	2130	15	2.7	23	1510	0.00048	5.3
	CV (%)	24	18	4	4	6	4	6	7	3	2	5	6	6	3	19	21	2
Organic 2	Mean	29	2.7	0.25	1370	156	0.11	0.054	17	27	19,200	2030	15	3.3	22	1400	0.00061	5.5
	CV (%)	30	22	4	6	6	6	7	4	7	1	2	4	9	2	14	30	6

(T). Concentrations of Co appear to be helpful in the separation of all three categories C, O, and T.

The box plots were remade excluding the soil-contaminated samples (Fig. 2). It can be seen that the plots for Br, Ca, Mn, and Rb are similar to those including the soil-contaminated samples (Fig. 1). The concentrations of Cs and K appear to help the discrimination of conventional coffee while Fe helps the discrimination of in-conversion samples, which was not possible when soil-contaminated samples were included (Fig. 1).

The analysis was further refined using a multivariate data visualization method considering an R scatterplot [15–17]. This is depicted in Fig. 3, for medium/long-lived nuclides, and Fig. 4, for short-lived nuclides, for samples from the categories organic (O) and conventional (C) with respect to each possible pair of elemental concentrations. The plots suggest that both categories of coffee samples can be well discriminated by pairs of elemental concentrations.

In the following step (Fig. 5), a statistical learning technique called classification tree was applied to summarize the information within the coffee database into an easily understandable representation for insights in class discrimination. Classification trees are often used in bot-

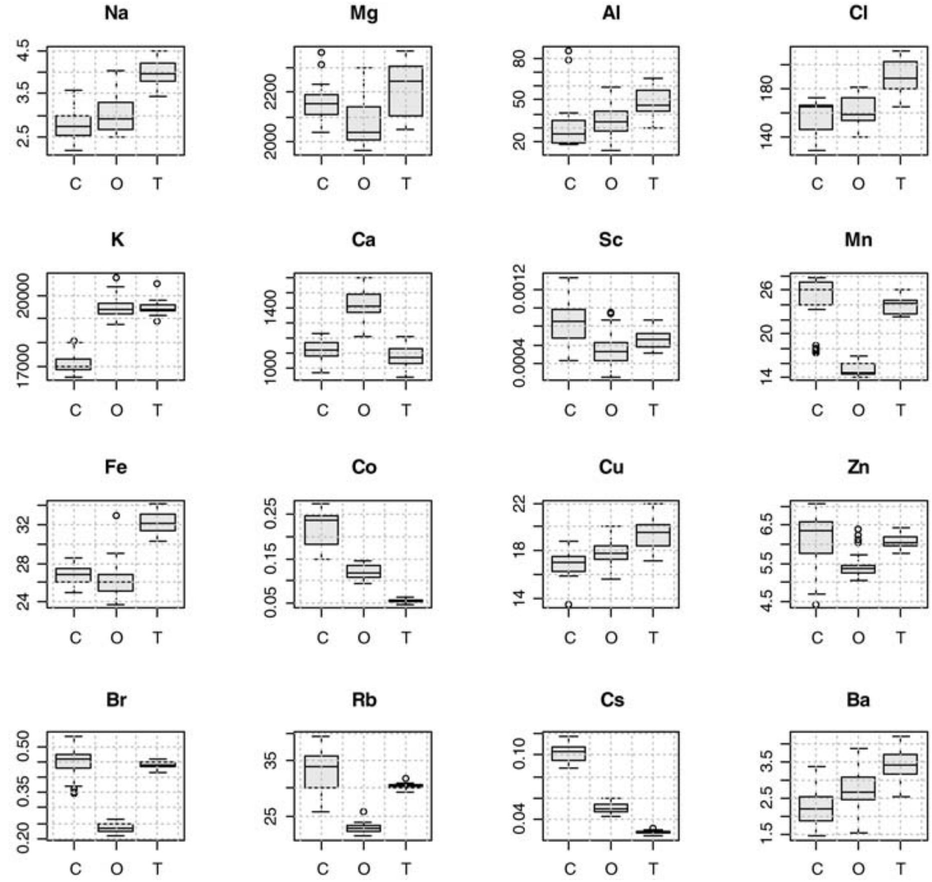


Fig. 2 Box plots of elemental concentrations (μg g^{-1}) for each category (without soil-contaminated samples in the conventional category)

any, medicine, entomology and artificial intelligence, but are less familiar to statisticians. The technique was implemented considering the Gini index as the criterion to guide the data split process. Additional details can be found in the specialized literature [17, 18].

The classification task was rather easy given the clear separation of categories already diagnosed by the direct visualization techniques. Figure 5 presents a classification tree in which all element concentrations available were considered for the tree building process. It resulted that only cobalt was needed for a complete classification. This tree reads as following: if the Co concentration is greater than or equal to 0.147 μg g^{-1}, go to the left and the sample is conventional (C), otherwise go to the right; if the Co concentration is greater than or equal to 0.079 μg g^{-1}, go to left and the sample is organic (O); otherwise the sample is from the in-conversion (T) category. The numbers depicted below the tree show the number of samples classified in each category. The indication 50/0/0 below the branch at the extreme left shows 50 samples classified by the rule in the category conventional, 0 in the category organic and 0 in the category in-conversion. Figure 6 shows a classification tree that excludes the possibility of using Co. In this case, Rb and Cs were the best elements selected for the classification (100% correct).

Finally, a Bayesian network was constructed using data-driven techniques to evaluate the information in the database. This technique takes into account the probabilistic framework introduced by Eqs. (1) and (2), being a solid foundation for expert systems aimed at diagnosis and discrimination. Details on this technique are presented elsewhere [19, 20]. The structure of the network considered here is very simple, assuming (in most cases) that the marginal probability distributions of elemental concentrations are conditionally independent, given the "true" category of the sample and knowledge on soil contamination. An arrow from one node (called parent node) to another node (called a child node) indicates that the values/categories of the child are probabilistic conditioned by the parent. Only four discrete categories of concentration values were defined for each element in this simplified network (associated with the quartile ranges observed in the database for each element concentration).

The computational implementation was developed with the software Netica (Norsys), which provides a convenient user interface for exploration and diagnosis us-

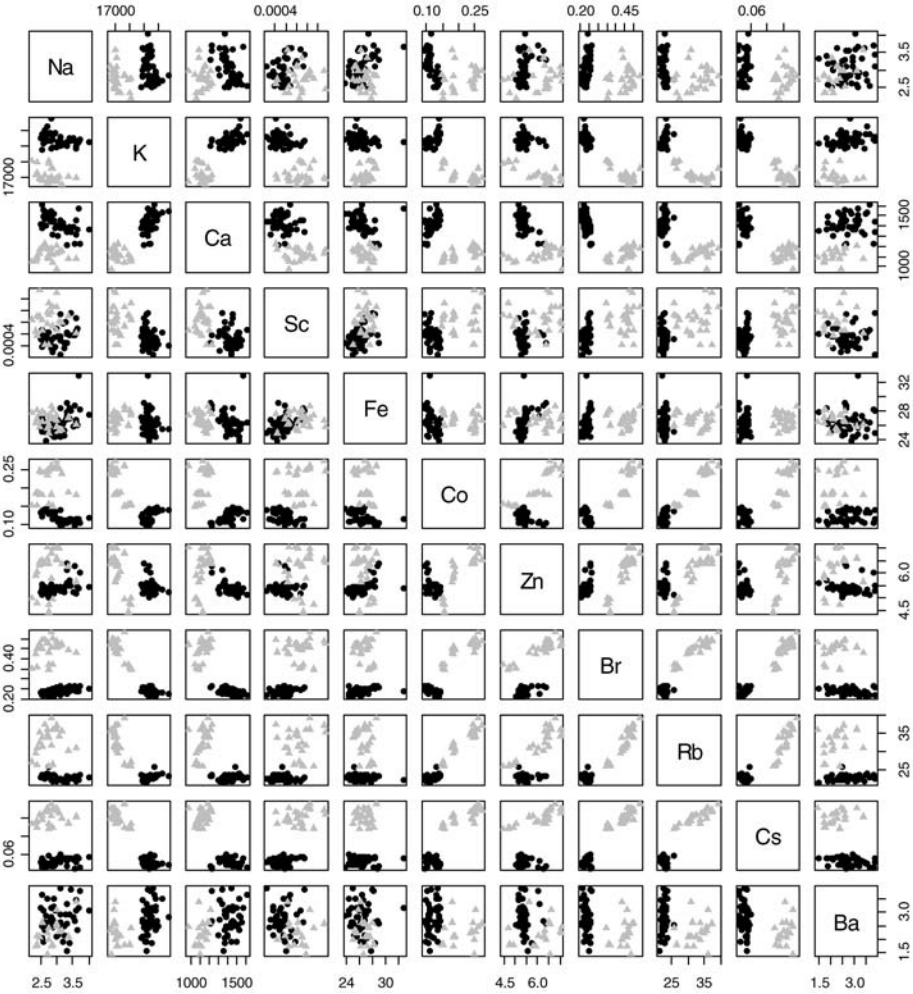


Fig. 3 Coffee samples (*black*, organic, *gray*, conventional) and pairs of elemental concentrations (μg g^{-1}) – w/o soil contamination samples (medium/long-lived nuclides)

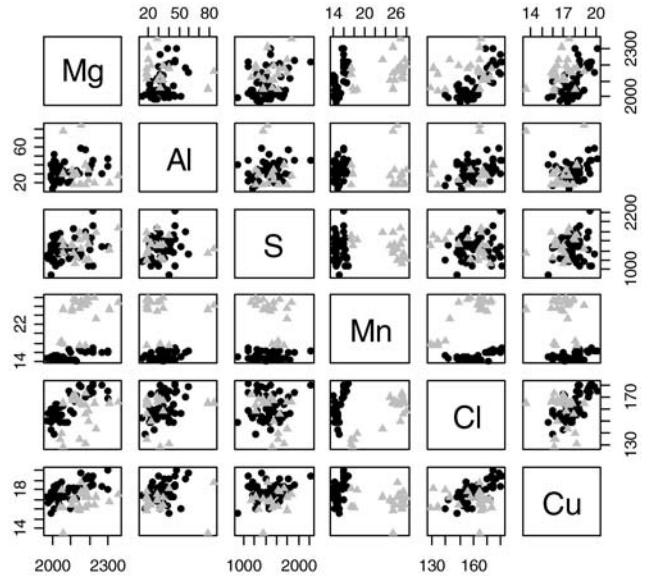


Fig. 4 Coffee samples (*black*, organic, *gray*, conventional) and pairs of elemental concentrations (μg g^{-1}) – w/o soil contamination (short-lived nuclides)

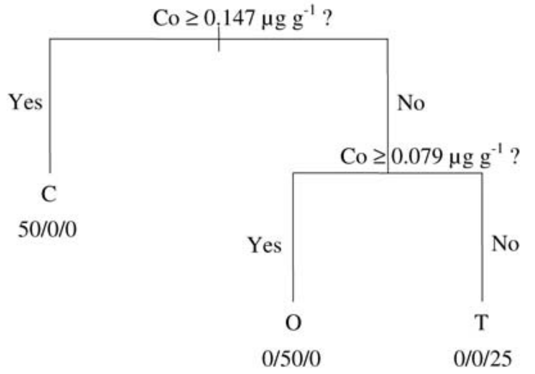


Fig. 5 Classification tree for soil-contaminated and noncontaminated samples (all elements allowed)

ing the network. Of 125 samples, 102 were randomly selected for the estimation of the conditional and unconditional probabilities in the network considering the frequency of cases in the database with the appropriate features. The network was adjusted in such a way that all possible events would have nonzero probability, to ac-

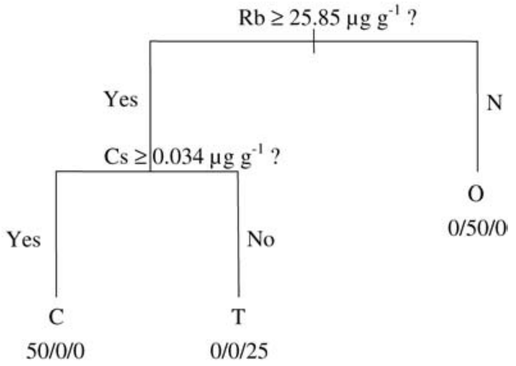


Fig. 6 Classification tree for soil-contaminated and noncontaminated samples (all elements allowed except Co)

count for *a priori* knowledge not reflected in the sample database. The conditional probability for each category was then computed by the system using Bayes rule.

Figures 7–10 show the conditional probabilities of each category of cultivation system (in the node Pmethod) and soil contamination (in the node Soil) given the available information on elemental concentrations measured in a certain sample.

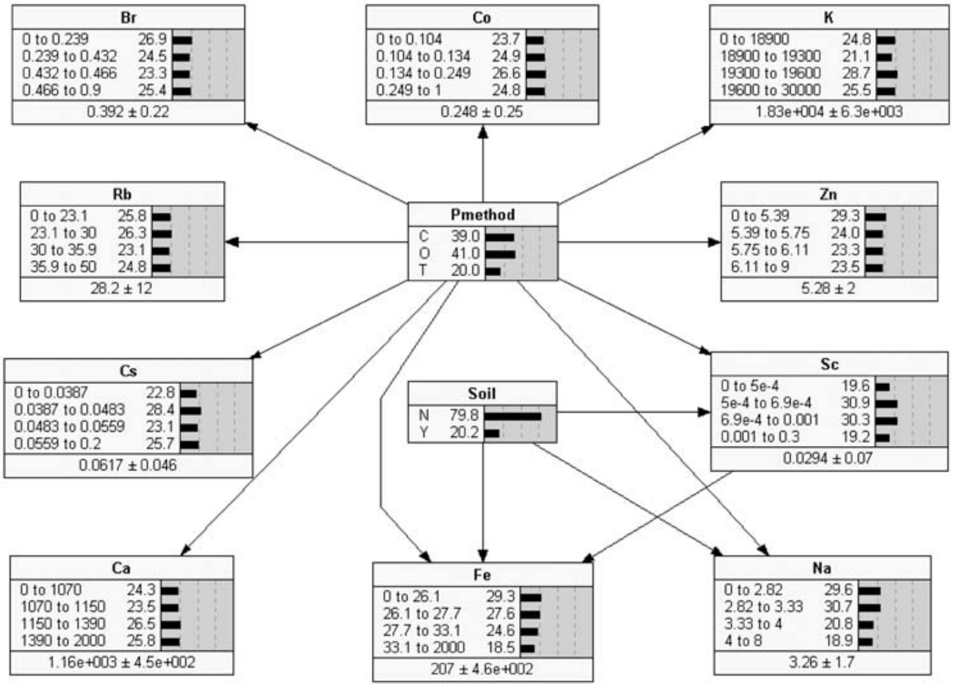
Figure 7 shows the probability of each category when there is no information on the elemental concentrations in a specific coffee sample. These probabilities are associated with the frequency of samples in the database within each category of cultivation system (C, O, T) and soil contamination (yes or no), in this implementation.

In Fig. 8, only the information on K, Ca, and Fe concentrations for a particular sample is assumed as known. The nodes associated with these elements are shown in a darker grey and the category of observed measurement set to probability 100 %. Elemental concentrations for other nodes (associated with other elements) are assumed as unknown. The probabilities shown in these nodes are the conditional probabilities for each category of elemental concentration for the sample given the observed concentrations for K, Ca, and Fe. The conditional probabilities for the cultivation system and soil contamination categories given the observed concentrations for K, Ca, and Fe are: 67.0 % for the sample being conventional, 31.7 % being organic, 1.38 % being transition (or in-conversion), and 91.6 % being noncontaminated by soil. The information on K, Ca, and Fe in this sample produced conditional probabilities for each category of cultivation system and soil contamination showing considerable uncertainty. Discrimination was not entirely satisfactory in such case.

In Fig. 9, only Br and Na concentrations are assumed to be known for the same sample. Now the conditional probabilities are 96.6 % for the sample being conventional, 2.74 % being organic, 0.7 % being transition, and 96.9 % being noncontaminated by soil. The discrimination was more satisfactory in this case.

In Fig. 10, it is assumed that all elemental concentrations are known for the sample. Thus, the conditional probabilities are close to 100 % for the sample being conventional, 0 % being organic, 0 % being transition and 98.6 % being noncontaminated by soil. Uncertainty

Fig. 7 Bayesian network without element concentrations set to specific levels (see text for details)



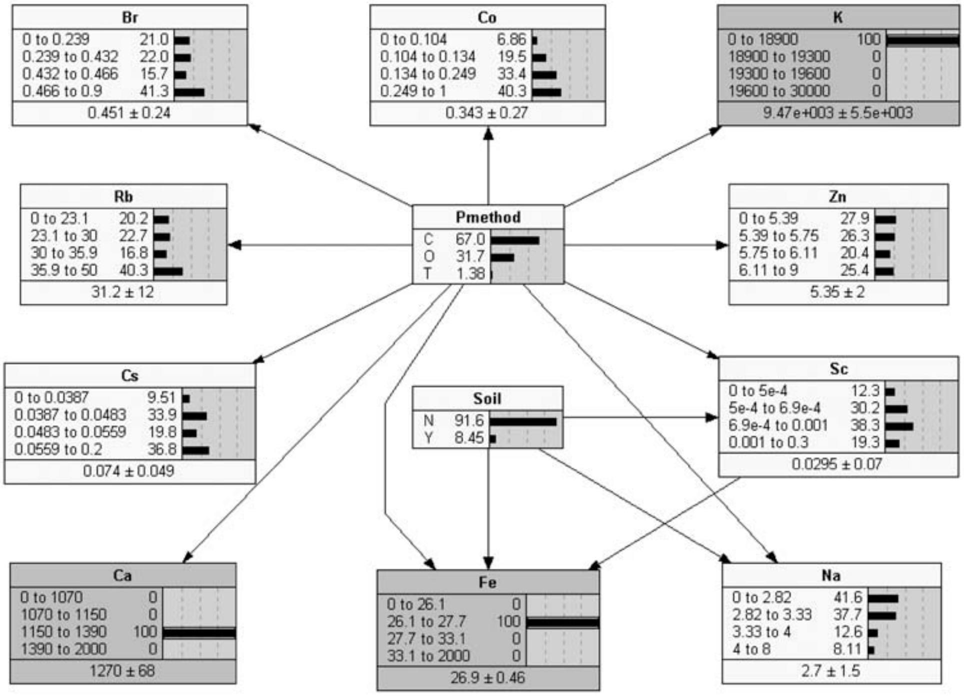


Fig. 8 Bayesian network with K, Ca, and Fe concentrations (μg g^{-1}) set to specific levels (see text for details)

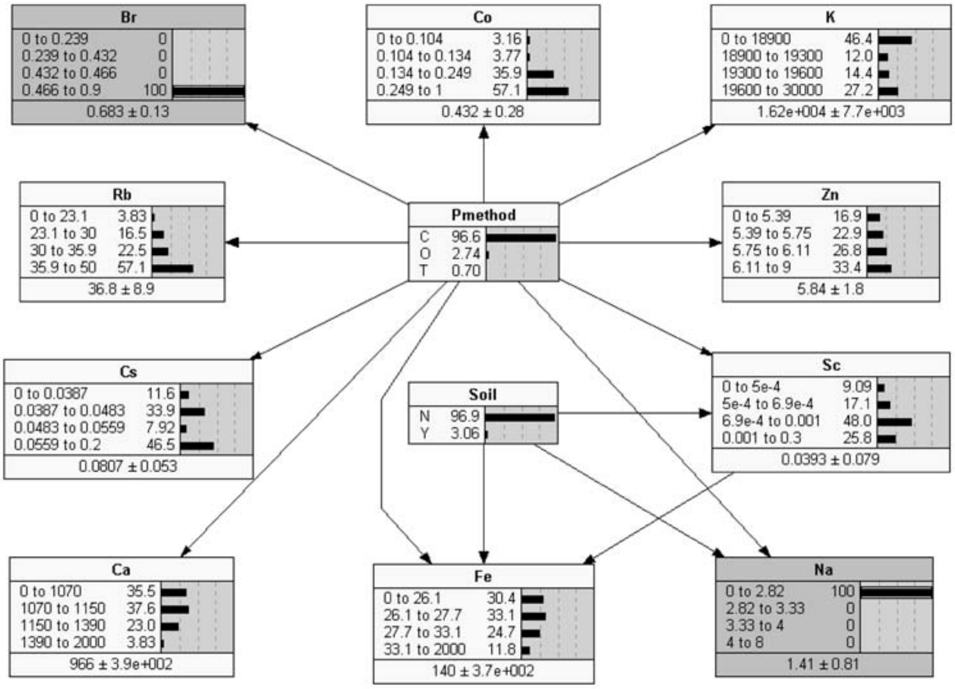


Fig. 9 Bayesian network with Br and Na concentrations (μg g^{-1}) set to specific levels (see text for details)

on the true category of cultivation system and soil contamination for the coffee sample was considerably reduced in this situation.

Finally, the discrimination accuracy of the network was tested with 23 samples not used in the network estimation (or learning) procedure. For each of these samples, all available elemental concentrations were informed to the network, in a situation similar to that indicated in Fig. 10. Table 2 shows the conditional probabilities for each category estimated by the network, given the elemental concentrations. Only the sample 10 was given a larger probability for the "wrong" cultivation

Fig. 10 Bayesian network with all element concentrations (μg g^{-1}) set to specific levels (see text for details)

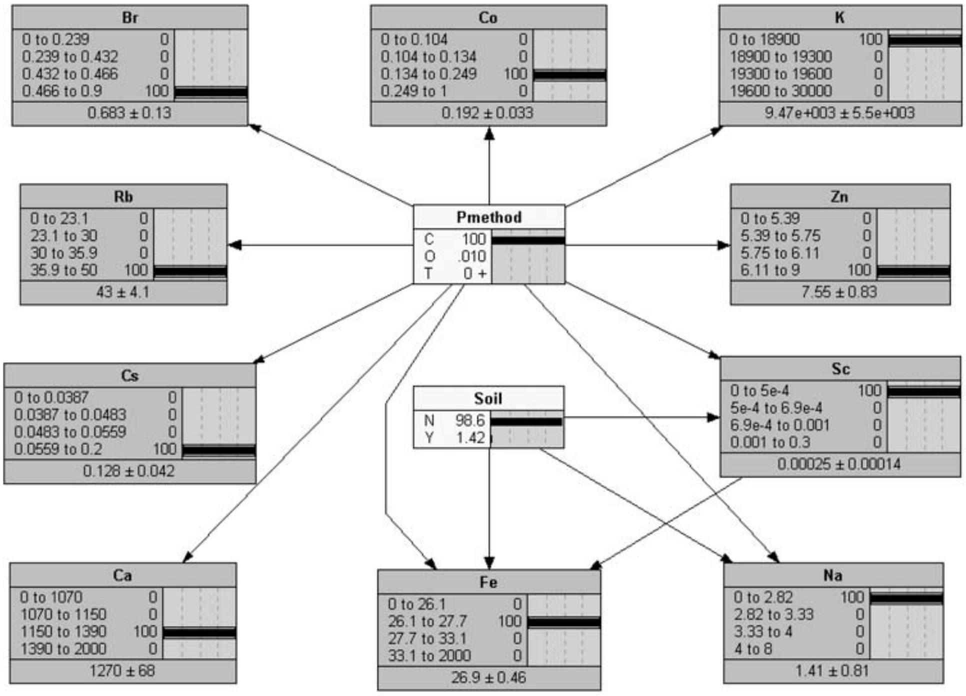


Table 2 Conditional probabilities estimated for a testing set of coffee samples not used to construct the Bayesian network (sample 10 was misclassified – right category is organic)

Sample number	Conditional probabilities (%) from Bayesian network, given the elemental concentrations					Sample category	
	Cultivation system			Soil contamination			
	C	O	T	Yes	No		
1	0	0	100	2	98	T	No
2	0	0	100	5	95	T	No
3	0	0	100	2	98	T	No
4	0	0	100	3	97	T	No
5	0	0	100	9	91	T	No
6	0	100	0	7	93	O	No
7	0	100	0	5	95	O	No
8	0	100	0	10	90	O	No
9	0	100	0	12	88	O	No
10	0	13	87	6	94	O	No
11	0	100	0	40	60	O	No
12	0	100	0	20	80	O	No
13	0	100	0	8	92	O	No
14	100	0	0	99	1	C	Yes
15	100	0	0	99	1	C	Yes
16	100	0	0	99	1	C	Yes
17	100	0	0	99	1	C	Yes
18	100	0	0	99	1	C	Yes
19	100	0	0	0	100	C	No
20	100	0	0	0	100	C	No
21	99	1	0	0	99	C	No
22	100	0	0	0	99	C	No
23	100	0	0	0	99	C	No

system category (in this case Transition). Organic was the presumed category for this sample. For the other 22 samples the presumed cultivation system category was given the highest probability (often close to 100%). Samples 14–18, known to be soil-contaminated, were correctly discriminated in this category with probability higher than 99 %. Samples 8 and 11 suffered from missing information on some elemental concentrations but still the higher probability was given to the presumed categories.

Concluding remarks

Results indicated a positive perspective for the use of elemental concentrations for the discrimination of organic from conventional green coffees. Since the nature of the database used in this research includes coffees from only one production region and harvest, the conclusions cannot be extrapolated into a more general context and need to be interpreted with caution. At this point, effort is being made towards the construction of a comprehensive database with information from conventional and organic coffee samples from other regions, harvests and coffee species (Arabica vs. Robusta). An evaluation of the discrimination performed by elemental concentrations based on the information from this new comprehensive database might suggest additional steps towards the development of a reliable methodology for coffee quality demonstration, concerning the organic and conventional attributes. Finally, research aimed at clarifying the causal mechanism associated with the differences observed in elemental concentrations, for organic and conventional coffees, is strongly recommended.

Acknowledgments The authors wish to thank the financial support supplied by FAPESP and CNPq; and TUDelft for granting a research fellowship to Mr. Tagliaferro. Thanks are also given to MINASCOFFEE for supplying the coffee samples, and to an anonymous referee for valuable comments.

References

1. De Bièvre P, Taylor PDP (2000) Fresenius' J Anal Chem 368:567
2. Krivan V, Barth P, Morales AF (1993) Mikrochim Acta 110:217
3. Martín MJ, Pablos F, González AG (1998) Anal Chim Acta 358:177
4. Martín MJ, Pablos F, González AG (1999) Food Chem 66:365
5. Martín MJ, Pablos F, González AG (1998) Talanta 46:1259
6. Costa Freitas AM, Mosca AI (1999) Food Res Int 32:565
7. Bode P, Fernandes EAN, Greenberg RR (2000) J Radioanalyt Nucl Chem 245:109
8. Frawley W, Piatetsky-Shapiro G, Matheus C (1992) Knowledge discovery in databases: an overview. AI Magazine, Fall:213
9. Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (1996) Advances in knowledge discovery and data mining. AAAI Press/The MIT Press, Cambridge, MA
10. International Federation of Organic Agriculture Movements (IFOAM) (2002) Organic agriculture world-wide 2002 – statistics and future prospects. IFOAM, Tholey-Theley, Germany
11. International Federation of Organic Agriculture Movements (IFOAM) (2000) Basic standards for organic agriculture. IFOAM, Tholey-Theley, Germany
12. FAO/WHO Codex Alimentarius Commission (2001) Guidelines for the production, processing, labelling and marketing of organically produced foods. GL 32–1999, Rev. 1/2001. FAO/WHO, Rome
13. Bode P (2000) J Radioanalyt Nucl Chem 245:127
14. Bode P, van Dijk CP (1997) J Radioanalyt Nucl Chem 215:87
15. Ihaka R, Gentleman R (1996) J Comp Graph Stat 5:299
16. General information on the R project is found at the site www.r-project.org
17. Venables WN, Ripley BD (1999) Modern applied statistics with S-Plus, 3rd edn. Springer, New York
18. Ripley D (1996) Pattern recognition and neural networks. Cambridge University Press, Cambridge, UK
19. Neapolitan RE (1990) Probabilistic reasoning in expert systems, Wiley, New York
20. Jensen FV (1996) An introduction to Bayesian networks. Springer, New York